ITC 6001 INTRODUCTION TO BIG DATA   3 US CREDITS

The course addresses students who are new to Data Science and are interested in understanding Big Data management and analysis techniques. Some programming experience is necessary to successfully complete the course.

Along with the theoretical aspects, the course provides hands-on experience with programming language libraries for data management, relational and NOSQL databases, stream processing as well as a distributed processing system.

ITC 6002 EXPLORING AND ANALYZING DATA   3 US CREDITS

The course focuses on procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of statistics which apply to analyzing data. It also examines probabilistic inference for generative models of inference together with standard techniques in pattern recognition. Topics covered include: Descriptive and inferential statistics, sampling, probability spaces, queuing theory, stochastic processes, mathematical modeling, experimental design, parametric and non-parametric tests, regression, clustering, Markovian and Bayesian networks.

ITC 6003 APPLIED MACHINE LEARNING   3 US CREDITS

The course provides a broad introduction to the key ideas of machine learning, a rapidly growing field which resides at the intersection of computer science and statistics, and is concerned with finding useful patterns in data. Emphasis is given on intuition and practical examples, which covers a wide range of real life implementations, including personalized product recommendations and natural language processing.

ITC 6004 DATA VISUALIZATION   3 US CREDITS

The course combines the science of data visualization with the art of graphic design, and introduces ways to accurately and effectively communicate complex information. Students are exposed to techniques of presenting complex ideas in easily accessible and understandable manners, by transforming data into visual graphics, such as charts, bar graphs, scatterplots, and heatmaps. The course also offers hands-on experience through exercises, which allow students to explore the types of data in use today, learn how people perceive different graphical displays, and create visual presentations that foster impact on the audiences.

ITC 6005 KNOWLEDGE GRAPHS   3 US CREDITS

The course provides a broad introduction to core ideas of the Semantic Web layer, an intellectual milestone for the evolution of the web and the management of data from heterogeneous sources, based on "The Web of Meaning" vision, which sets the basis for significant developments in technical prerequisites and business requirements. Students are exposed to a multidimensional overview of the constitutional elements of the semantic web, namely: RDF, OWL, Query, Logic, Proof, Trust. The emphasis is on the integration of core semantic web ideas to real world problems and on the capacity of students to vision real world application by adopting ontological engineering and extensive reasoning capabilities in data.

ITC 6107 BIG DATA ARCHITECTURES   3 US CREDITS

This course prepares students to deal with large-scale collections of data as objects to be stored, searched over, selected, and transformed for use. Emphasis is placed both on the background theory and the practical application of information retrieval, as well as, database design and management, data extraction, transformation and loading for data warehouses, and operational applications.

ITC 6008 SEARCH ENGINES AND WEB MINING   3 US CREDITS

This course provides a comprehensive introduction to the theory and implementation of algorithms for organizing and searching large text collections. The first half of the course examines text search engines for enterprise and web environments; an open-source engine such as Indri can be used as a working example. The second half of the course explores text mining techniques such as recommender systems, clustering, and categorization. The course strives for a hands-on experience in document ranking, evaluation, and classification into browsing hierarchies, as well as other related topics.

ITC 6009 MACHINE VISION IN DATA SCIENCE   3 US CREDITS

This course emphasizes problem segmentation in the field of machine vision. The goal is to master processing techniques and enlighten machine learning techniques on vision problems, to inform about new developments on learning, and to identify unique challenges and opportunities in the combined fields.

Computer Vision has become ubiquitous in our society, with applications in search, image understanding, apps, mapping, medicine, drones, self-driving cars etc. Core to many of these applications are visual recognition tasks such as image classification, localization and detection. The explosion of images and videos on the Internet and the availability of large amounts of annotated data have created unprecedented opportunities and fundamental challenges on scaling up computer vision.

ITC 6010 NATURAL LANGUAGE PROCESSING   3 US CREDITS

The course explores fundamental concepts and ideas in natural language processing (NLP), otherwise known as computational linguistics, and aims to develop an in-depth understanding of both algorithms for processing linguistic information and the underlying computational properties of natural languages. Students are exposed to word-level, syntactic, and semantic processing from both a linguistic and an algorithmic perspective, led by current research and tools in the field. The course focuses on modern quantitative techniques in NLP -- using large corpora, statistical models for acquisition, disambiguation, and parsing -- and the construction of representative systems. Topics to cover are: natural language understanding, morphological processing, lexicon, tagger, converter, parser, word sense disambiguation, word and sentence embeddings, deductive approaches to interpretation, machine translation and language acquisition.

ITC 6015 STRATEGIC THINKING FOR DATA SCIENTISTS   **3 US CREDITS**

This course examines the science of strategic management equipping students with the business knowledge they need to harness the possibilities of data analytics and leverage data analytics for effective decision making. An understanding of business objectives and the underlying processes that drive functional operations are emphasized as well as the development of crucial soft skills such as communication and teamwork skills that will help data scientists generate business value for their companies. Topics include, although not exclusively, the understanding of business strategy, how it is formed and the differences between various strategic options.

The formation of competitive advantage through functional-level strategies will be discussed together with frameworks of developing data-analytics strategies and data-driven organizations. At the same time, in order to navigate the challenges data scientists face in business environments students will develop a number of soft skills, such as creativity, cognitive flexibility and empathy and the ability to work effectively in multi-functional teams. In addition, students will learn how to communicate the results of their analysis to key decision-makers and demonstrate how they will help to improve performance. Central within this module is the development of students' understanding of how businesses make decisions in order to direct their data analytic efforts in the right direction and deliver the kinds of insights that are highly valuable and actionable. A powerful "Business Simulation" will be used throughout the course to deepen student's understanding of the real factors which impact business decisions and practice what they learn during the course.

ITC 6420   SAS PLATFORM FOR BUSINESS ANALYTICS   **3 US CREDITS**

Prerequisites:

ITC 6001 INTRODUCTION TO BIG DATA
ITC 6002 EXPLORING AND ANALYZING DATA
ITC 6004 DATA VISUALIZATION
ITC 6003 APPLIED MACHINE LEARNING

In today's business environment every organization is flooded with data, scattered around all departments in various formats. These data are related to sales and demand, customers, suppliers, personnel, financial information, inventories etc. In order for organizations to survive in today's volatile and uncertain environment they must have competitive advantages that can be acquired by exploiting the large amounts of data that they have so as to become more efficient and more effective in their decision making processes. This course focuses on the three areas of data exploitation for decision making purposes i.e. data management, business intelligence and business analytics. The software to be used during the course to support the above three areas and related concepts is SAS Enterprise Guide, SAS Enterprise Miner and SAS Visual Analytics.

In addition, the course will form the foundation and will prepare students to participate in the SAS Joint Certificate in Business Analytics and Data Mining that will be provided by SAS after the successful completion of a separate examination process (separate project, exam and presentation of outcomes).

ITC 6230 DEEP LEARNING   **3 US CREDITS**

Prerequisites:

ITC 6002 EXPLORING AND ANALYZING DATA
ITC 6003 APPLIED MACHINE LEARNING

The course focuses on Deep Neural Network Cell architectures. Deep learning is a branch of machine learning concerned with the development and application of modern neural networks. Deep learning algorithms using modern NN architectures extract layered high-level representations of data in a way that maximizes performance on a given task. We will cover their theory, their applications, and their training.

In particular, the course provides a background on NN architectures starting with Feed Forward and Recursive Neural Networks (FFNN, RNN) and the derivation of Backpropagation as an organized computation of Gradient Descent for their training from data. We address limitations stemming from the vanishing and exploding gradients problems. Then we move to more complex architectures: Convolutional Neural Networks (CNN), a variety of Sequence modeling NN like Long Short Term Memory NN (LSTM), Gated Recurrent Units (GRU), Transformers, Reformers. We also discuss Attention mechanisms to increase long term learning in Sequence NN's. We revisit Back Propagation and explain how the latest NN cell architectures solve or mitigate the vanishing/exploding gradients problem of traditional RNN's. We finally cover optimization techniques adapted to NN training, hyperparameter tuning and explore Python libraries like Pytorch and Tensorflow for efficiently program and train NN's.

ITC 6125 MACHINE LEARNING AND APPLICATIONS 3 US CREDITS
Prerequisites:
ITC 6003 APPLIED MACHINE LEARNING

Machine learning algorithms are data analysis methods which search data sets for patterns and characteristic structures. Typical tasks are the classification of data, automatic regression and unsupervised model fitting. The course will cover methods from a variety of related subjects including statistics, applied mathematics and more specialized fields, such as pattern recognition and neural computation. Those methods will cover image and speech analysis, medical imaging, bioinformatics and exploratory data analysis.

ITC 6421 CAPSTONE PROJECT 3 US CREDITS
Prerequisites:
ITC 6107 BIG DATA ARCHITECTURES
ITC 6008 SEARCH ENGINES AND WEB MINING
ITC 6420 SAS PLATFORM FOR BUSINESS ANALYTICS OR ITC 6230 DEEP LEARNING
ITC 6009 MACHINE VISION IN DATA SCIENCE

The capstone project is undertaken by students working individually or collaboration and in consultation with an instructor who acts as their capstone advisor. The capstone project must convey a sound solution to a practical problem. The capstone project is a significant analytical, design and implementation piece work and provides an opportunity for students to draw on their methodological, analytical and substantive learning in a comprehensive written study in the field of data mining and big data.

ITC 6440 THESIS 3 US CREDITS
Prerequisites:
ITC 6008 SEARCH ENGINES AND WEB MINING
ITC 6230 DEEP LEARNING
ITC 6107 BIG DATA ARCHITECTURES
ITC 6009 MACHINE VISION IN DATA SCIENCE

The thesis is undertaken by students working individually and in consultation with an instructor who acts as their thesis advisor. The thesis must convey scholarly and/or professional analysis informed by appropriate application of methodology. The thesis is a significant analytical piece of work and provides an opportunity for students to draw on their methodological, analytical and substantive learning in a comprehensive written study in the field of data mining and big data.